CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

CiFE - Citation Function Extraction from Citations' Context

by

Taimoor Riaz

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Computing

Department of Computer Science

2020

My dissertation work is devoted to My Family, My Teachers and My Friends. I have a special feeling of gratitude for My beloved parents, brothers. Special thanks to my supervisor whose uncountable confidence enabled me to reach this milestone.

CERTIFICATE OF APPROVAL

CiFE - Citation Function Extraction from Citations' Context

by

Taimoor Riaz

(MCS181044)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Muhammad Arshad Islam	NUCES, Islamabad
(b)	Internal Examiner	Dr. Azhar Mahmood	CUST, Islamabad
(c)	Supervisor	Dr. Muhammad Abdul Qadir	CUST, Islamabad

Dr. Muhammad Abdul Qadir
Thesis Supervisor
December, 2020

Dr. Nayyer Masood
Head
Dept. of Computer Science
December, 2020

Dr. Muhammad Abdul Qadir
Dean
Faculty of Computing
December, 2020

Author's Declaration

I, **Taimoor Riaz** hereby state that my MS thesis titled "**CiFE – Citation Function Extraction from Citations' Context**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Taimoor Riaz)

Registration No: MCS181044

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**CiFE – Citation Function Extraction from Citations' Context**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Taimoor Riaz)

Registration No: MCS181044

Acknowledgements

Firstly, I am very grateful to Allah the Almighty who gave me this opportunity and courage to complete my MS thesis.
Then, I am very thankful to my supervisor Dr. M. Abdul Qadir who helped me out throughout my whole MS Thesis and his kind guidance and support helped me a lot in completing my MS Thesis. I am so grateful to him that he shared his pearl of wisdom and his expertise greatly assisted me in my Research Thesis.
I am also thankful to all the people who helped me in my thesis.
Finally, I would like to thank my parents for their support and guidance as without their support and encouragement it would not be possible for me to complete my MS thesis.

(Taimoor Riaz)
Registration No: MCS181044

Abstract

Automated classification of citations' functions in scientific text is a new emerging research topic inspired by traditional citation analysis in bibliometric fields. The aim is to classify citations in scholarly publication in order to identify author's purpose or motivation for citing a particular paper. For this purpose, several citation schemes have been proposed to classify the citations into different functions accurately. However, it is a challenge to extract functions from citations' context with high recall. To address the challenge, this thesis adopts eight citations' functions taken from CCRO classes to develop a machine learning system to maximize recall without compromising the precision. This machine learning system could be utilized in bibliometric applications for categorizing the links between the citing and cited papers into eight citations' functions, which can then be used to build a meaningful knowledge graph for the published research papers. For this purpose, we have conducted a survey of the available citations' functions and citations' functions classifiers. Afterwards, we adopted a minimum set of eight citations' functions with minimum overlapped meanings and also best machine learning methods (SVM, NB, RF) have been selected for extraction of citations' functions from citations' context. Athar's data set have been used which is annotated in eight citations' functions. Several types of features that capture the characteristics of citation sentences are extracted by devised feature extraction rules are served as the inputs of automatic classifiers. A data set have been built using the proposed scheme and a number of experiments have been carried out to assess the model. 98% weighted - average F1-Score and 90% macro F have been achieved. Experimental results have shown that the proposed approach outperforms the existing methods in terms of Macro precision, Macro recall and Macro F. This classifier is useable in digital libraries to categorize the cited articles into eight citations' functions accurately. The categorization of cited paper in eight citations' functions facilitates the researcher to get understanding of cited paper even before and without reading that paper. With the help of this proposed system, scholarly community will be able to find maximum number of relevant research papers within minimum time span unlike traditional methods.

Contents

List of Figures

List of Tables

Abbreviations

ACL	Association for Computational Linguistic
ADVMOD	Adverbial Modifier
ADJ	Adjective
ADP	Adposition
ADV	Adverb
CCRO	Citation's Context and Reasons Ontology
CONJ	Conjunction
ML	Machine Learning
NB	Naïve Bayes
NSUB	Nominal Subject
PREP	Preposition
POS	Part of Speech
RF	Random Forest
SVM	Support Vector Machine
SCONJ	Subordinating Conjunction

Chapter 1

Introduction

1.1 Background

Digital libraries contain a vast number of research publications online. These libraries enable researchers to find information about the publications and their citations. A citation is a reference or link between the current research work and the previous relevant studies. Citation links build towards a large graph of knowledge and show many useful information. A very simple information is the number of citations of a particular paper which are used to compute the reputation of a researcher and publication venue. Impact factor and different indices are computed from the number of citations which are being used for research ranks and productivity. Many appointments and promotions are being decided based upon the citations. Counting of number of citations is not sufficient as the reason for which a paper is being referred by an author is multifold. A paper is cited to give a research background in the topics, to indorse or disagree with the previous research, to see the emergence of a particular area of study, to enhance a particular concept, to critically investigate the precious work and many more. However, the reason for which a paper is being cited (citation function) is hidden in the context of the citation text in the citing paper.

The importance of an article is commonly determined by the number of times it is cited. In this way the initial research of an analysis of citations were, however, primarily concerned with citing frequency and other citing statistics, and therefore all citations were equally treated [1]. Consequently, there was a major flaw in these techniques of citation analysis as they did not differentiate between the negative and positive citations. However, with the rapid development of information technology, more and more academic publications are accessible in electronic format and convenient for users to access. The previous studies of citation analysis do not seem as effective and accurate to retrieve the desired articles on the basis of the citing frequency and standard citation indexes. In addition, having lots of documents retrieved will cost readers a lot of time to find their desired information they expect. For these reasons, the research papers should be filtered by citations' reasons (citation function) and linked to articles which satisfy the needs of a reader by creating a citation-based network. Depending upon the context of citation, the citations is going to convey a very useful information to the scholar community to build a meaningful knowledge graph which can be queried. Therefore, discovery of correct citations' functions is an important task being handled by the research community in scientometric or bibliometric analysis. Discovery of citations' functions is being termed as citation classification.

Since five decades, a series of classification schemes for citations were formulated and developed. Garfield [2] in his research, has listed 15 reasons for citing other people's work and provided a valid signal that the citation and its nature is important. His work paved the way for the analysis of citations' functions. Later on, Lipetz established a 4-category scheme defining the relationships between the cited and the citing articles [3]. Afterwards, Athar proposed sentiment analysis of scientific citations [4]. He classified the citations into three sentiment classes Positive, Negative and Neutral as based on the sentimental context of citations. Moreover, these sentiment classes were too simple to cover the diverse reasons of Citations' functions in the citation context. In order to improve the efficiency of the classification of large amount of data using computer technology, a number of researchers are working to develop a scheme that could be easily adapted by the automatic classifiers.

A detailed study shows that the literature contains more than 150 reasons for defining citations relationship among articles. A research was conducted in 2014, which revealed that the size of data to be processed for the citation graph was estimated that Microsoft Scientific Research had more than 45 Million research papers, 55 million on the Web of Science and 100 million on Google Scholar [5]. Annotating every citation into 150 reasons is nearly impossible. In addition, the reasons gathered for the citation have overlapped and diffused meanings. To discover these reasons for the citations by using machine learning algorithms, will result in low accuracy. Thus, it is necessary to automatically discover the reasons for the citations. To achieve this goal, a researcher developed a minimal set of eight citations' functions which are disjointed in nature and are formally represented in the form of Ontology [6]. This minimum set of reasons defines Citation's Context and Reasons Ontology (CCRO) classes, which allow machine learning algorithms to identify the reasons for these citations accurately.

1.2 Problem Statement

It was observed that the precision (accuracy) and recall (completeness) of the method to extract citations functions (reasons) from citations' context needs to be improved. The improvement can be achieved by addressing at least two issues that are to come up with a more concrete definition of the citations' function and then develop machine learning system to maximize precision and recall. This research addresses the second aspect that is to develop a machine learning system for concrete definition of the citations' functions and compare it with the existing state-of-the-art systems.

1.3 Research Questions

To solve the problem, as indicated in the problem statement section, we need to address at least following questions:

1. What are concrete classes of the citations' functions already published in the literature?
2. What are the best machine learning methods to classify citations' context into citations' functions?
3. What are the deficiencies in the classification systems?
4. How can a classification system be developed by addressing the gaps in the existing systems?

1.4 Purpose

The goal of this study is to classify the citations into CCRO classes by using supervised machine learning algorithm. In this way, we will be able to conclude that CCRO classes are most appropriate to citation texts as they hold a significant potential to improve performance of citations' functions classification.

1.5 Scope

In this research, we are going to formulate and develop a system with improved precision and recall for the identification of classification functions as defined by recently proposed citation ontology, CCRO [6].

1.6 Significance

Citations' functions classifier could be utilized in bibliometric applications for categorizing the links between the citing and cited papers into CCRO classes, which can then be used to build a meaningful knowledge graph for the published research papers. The knowledge graph can then be used to answer meaningful queries related with the citations' functions as an application in modern digital libraries.

1.7 Methodology

The proposed methodology to develop a system for the extraction of citations' functions from the citations' context is described. As given in the previous chapter that there are four major questions to be answered in developing a system with high precision (accuracy) and recall (completeness). Methodology ton answer each question is described in the following sections:

1.7.1 Selection of Concrete Citations' Functions

Concrete citations' functions mean a minimal set of citations' reasons with minimum overlapped meanings. A detailed critical survey of the published citations' reasons is required in order to answer the first questions. In the survey, English Language dictionary will be used to judge the overlapped meanings of the reasons. Similar work will also be looked in detail. The results of this activity is going to be a minimum set of citations' reasons with minimum overlapped meanings, which have been termed as concrete set of citations' functions.

1.7.2 Selection of Best Methods to Classify Citations' Context in to Citations' Functions

Most of the techniques proposed to resolve the citations' functions classification issues are applied by a classifier from the field of machine learning, which is trained on different features of citation context. We will investigate the best machine learning methods to classify the citation texts into citations' functions. Thus, considering the aspects of an automated system for classification of citations, we will explore various types of features from the existing state-of-the-art approaches. In the citation function classification, the selection of features is an important technique. These features include syntactic, semantic and citation specific features. We will extract all the important features and convert all the features into feature vectors for numerical representation which is required for an input to machine learning (ML) classifiers for testing and training. We will analyze the ML approaches and compare their strengths and weaknesses. On the basis of their strengths and weaknesses we will select the best machine learning method to classify citation texts into Citations' Functions accurately.

1.7.3 Identification of Deficiencies in the Classification Systems

As we have discussed above that we will conduct a survey and look for deficiencies of existing classification systems. First of all, we will highlight the issues of overlapped and diffused meanings of citations' functions classification schemes. Afterwards, we will be able to select a disjointed set of citations' functions. This disjointed set of citations' functions will allow machine learning classifiers to identify the citations' functions from citation texts accurately. Secondly, considering the high recall of an automatic framework for citations' functions classification, we will study several types of features from citation texts. These features include syntactic, semantic and citation specific features. Thus, on the basis of deficiencies and weaknesses in these features, we will devise feature selection rules for extraction of important features from citation texts to improve the recall of citations' functions classification.

1.7.4 Implementation of a Classification System

In this section, we are going to develop a machine learning system to extract the citations' functions from the citation texts. In order to do this, we will use the best machine learning method to classify citation texts into Citations' Functions accurately. The machine learning method require annotated data set. For this purpose, we will select data set of citation text. Then we will divide it into two sets; one smaller and the other larger set. The whole process will be completed in two phases. In the first phase, we will annotate the small set of data set. This annotated data set will help to train the machine learning classifier to classify the citation texts into citations' functions. After annotation process, we will devise feature selection rules to extract important features from citation texts to improve the accuracy of Citations' functions extraction. Afterwards, in the second phase we will divide the larger unannotated data set into multiple subsets and follow an iterative process to annotate it with combination of machine learning and manual verification. The classifier will be trained on annotated data set and it will predict the functions for first subset of citation texts from the unannotated data. This predicted subset will be made a part of training data set after manual verification of 10% machine predicted data. Then the second subset will be predicted in the same way and will be made a part of training data set. Likewise, all the subsets of larger data set will be predicted in the same way. After the completion of prediction process, we will pick 15% citation texts randomly from predicted data and it will be manually verified. We will adopt k-fold cross validation technique to evaluate the classification results of citations extraction system. In this way, we will develop a large data set to train a model to classify a citation from the citation context into citations' functions.

Chapter 2

Citations' Functions

From the past several decades, the classification of citations' functions have been researched and studied extensively. Scholars from different domains of science suggested various methods to examine and explain the complexities between the citations' functions.

2.1 Classes of the Citations' Functions

In 1960s, semantic analysis started, eventually becoming the dominant technique in citation content studies. For the purpose of citation motivation the researcher Garfield [2] was the first one to suggest further investigation. For automatic computation of a citation classification, he presented 15 different scenarios (reasons) in which the author cite someone other's research work and provide a valid signal that the citation and its nature is important. Latterly, most researchers used these reasons for the identification of semantic citation characteristics. Moreover, with these 15 reasons the 4 categories scheme, namely: scientific contribution, continuity relationship, disposition of contribution and non-scientific contribution of Lipetz [3] on relationships between citing and cited articles are also used by the researchers. These four citations' functions have no crystal clear definitions, moreover these citations' functions have maximum overlapped boundaries.

The studies of citation classification are usually based on four-dimensional citation schemes which have presented by Moravcsik and Murugesan [7]. They distinguished between confirmative, conceptual, evolutionary and perfunctory. The count of presented categories (also called citation functions) ranges between 3 and 35 [2][8][9]. One of the research study [8] utilized the POS tags for identification of grammatical subjects and further categorized as different agent forms, whereas Mercer and DiMarco demonstrated the utility of rhetorical references in citation classification [9]. Both studies indicate the possible ability to distinguish citations by syntactic features.

S.Teufel also studied well the automated method and a scheme for classification of citations' functions [8]. Throughout early works of Teufel considered the varied author writing styles from various domain relevant to different parts of the article and divided the sentences in which the statements of authors appeared into twelve categories. The approach has compared many citation feature schemes from the last century, arguing that most of them are too sociologically focused and therefore difficult to operate without expert knowledge of sociology and apply in other fields. In the scheme of twelve classes, the Pbas, Puse and Pmodi have defused meanings. For example if there is a citation text, we get confused as the author uses cited work as basis or starting Point. On the other hand, PMot and PSup are also overlapped. As PMot is used to motivate work in current paper while PSup provide support for each other. Thus, most of the classes from these twelve classes are overlapped.

Dong and Schäfer [10] employed a learning based method with multiple features like negation, cue words, POS-tag and position. Their main focus on perfunctory dimension for citations in Moravcsik and Murugesan [7] schema and categorized the citations into four general categories such as 1) background, 2) technical basis, 3) fundamental idea and 4) comparison. For evaluation, they have used 120 papers which are extracted from ACL Anthology and achieved 0.66 macro-F. Moreover, their research work integrate the explicit and implicit citations for the purpose of classifying citation's functions. In this scheme of citations' functions, the three classes namely; background, technical basis and fundamental idea are overlapped as well as defused meanings to some extent.

Afterwards, Athar proposed sentiment analysis of scientific citation [4]. He classified the citations into three sentiment classes such as 1) 'Positive', 2)'Negative' and 3) 'Neutral', based on the sentimental context of citations. They have employed some set of feature (structure-based) for the purpose of training ML classifier. Moreover, they have used the citing sentence for predicting sentiments. But it was too simplified, moreover the negative and positive classes are overlapped with neutral class to some extent. Afterwards, Butt et al. [11] used NB classification technique to classify citation sentence into positive and negative sentiments. Moreover, they have employed syntactic based features but did not attempt to utilize the semantic based features which helps to recognize the authors' sentiments.

The research study of Abu-Jbara et al. [12] developed a classification scheme of six category which was mostly selected from Teufel's 12 categories [8] for to better serve bibliometric measures and applications. These six categories are criticism, comparison, use, substantiation, basis and neutral. For experimental purpose they employed 3271 citations which were gathered from ACL Anthology. Moreover, these citation were annotated on the basis of their polarity and purpose. It seems that three categories substantiation, neutral, use and basis are overlapped. Sometimes it becomes impossible to differentiate between these four overlapping classes. Along with this, the remaining two classes' criticism and comparison are also overlap with neutral class.

Imran developed a minimal set of citation's context and its reasons which are disjointed in nature and are formally represented in the form of Ontology [6]. They have reduced more than 150 reasons in to 8 Citations' Functions. These citations' functions are named as Incorporate, Based On, Extend, Negate, Criticize, Contrast, Compare and Neutral. Each citations' function represent a unique citation link among the research article. This minimum set of citations' functions defines the Citation's Context and Reasons Ontology (CCRO) classes. The researcher identified and extracted the dominant verbs in a citation text by using NLP techniques. Afterward, he mapped citation texts into CCRO classes with the help of dominant verbs. Survey of the citations' functions are shown in Table 2.1.

Table 2.1: Survey of the Citations' Functions

Sr #	**Scheme**	**Number of Functions**	**Citations' Functions**
1	Liptez [3] 1965	4	Scientific Contribution Continuity Relationship Disposition of Contribution Non-Scientific Contribution
2	Moravcsik and Murugesan [7] 1975	4	Conceptual Evolutionary confirmative Perfunctory
3	Tuefel, Siddharthan, Tidhar [8] 2006	12	Week CoCoGM CoCo CoCoRo CoCoXY PBas PUse PModi PMot PSim PSup Neut
4	Dong Schafer [10] 2011	4	Fundamental idea Technical basis Back ground Comparison

Table 2.1- Continued from Previous Page

Sr #	Scheme	Number of Functions	Citations' Functions
5	Athar Teufel [13] 2012	3	Positive Negative Neutral
6	Abu Jbara, Ezra, Radev [12] 2017	6	use Substantiating Basis Criticize Comparison Neutral
7	Imran Ihsan Abdul Qadir [6] 2019	8	Incorporate Extend Based On Negate Criticize Contrast Compare Discuss

2.2 Selection of the Minimal Set of Citations' Functions with Disjointness

We studied the literature comprehensively and found that there are multiple citations' functions for identifying a citation relationship between articles. As Imran analysis shows that the literature contains more than 150 citations' functions for identifying a citation relationship between articles [6]. Imran has claimed in his research, these citations' functions for the citation texts have overlapped as well as diffused meanings and annotating every citation into 150 citations' functions is nearly impossible. Furthermore, to discover these functions for the citations by using ML algorithms, will result in low accuracy. In this way, we have adopted a disjointed minimal set of 8 citations' functions proposed by Imran [6] as he addressed all the problems of overlapped as well as diffused meanings of citations' functions classification schemes. These 8 citations' functions are shown in Table 2.2.

Table 2.2: Citations' Functions

Context Class	**Citations' Functions**	**Collaborative Meaning**
Positive	incorporate	To cite a research as part of a whole
	Extend	To spread from a central research to a wider solution
	Based On	To use a research as foundation or starting point
Negative	Negate	To cause to be ineffective or invalid
	Criticize	To find fault in a research with: points out the faults of
	Contrast	To show differences with opposite nature
Neutral	Compare	To examine in order to show similarities
	Discuss	To consider or examine by argument

2.3 Examples of Citations' Functions with Citation Texts

The examples of the sentences used for a particular function in the citation context are illustrated below. All these examples are given along with their citations' functions in Table 2.3. In this table, the first column represent sentiment classes. The second column represent CCRO classes and third column represent examples.

Table 2.3: Examples of Citations' Functions with Citation Texts

Context Class	**Citations' functions**	**Examples**
Positive	Incorporate	Smith and Smith (2007) describe a more efficient algorithm that can compute all edge expectations in O(n3) time using the inverse of the Kirchoff matrix K1.
	Extend	Stochastic models (Cutting et al., 1992; Dermatas et al., 1995; Brants, 2000) have been widely used in POS tagging for simplicity and language independence of the models.
	Based On	One of the most effective taggers based on a pure HMM is that developed at Xerox (Cutting et al. , 1992).
Negative	Negate	Therefore, sublanguage techniques such as Sager (1981) and Smadja (1993) do not work.
	Criticize	Chiang (2005) introduced a constituent feature to reward phrases that match a syntactic tree but did not yield significant improvement.
	Contrast	With all but two formats IBI-IG achieves better FZ=l rates than the best published result in (Ramshaw and Marcus, 1995).
Neutral	Compare	Actually, it is defined similarly to the translation model in SMT (Koehn et al., 2003).
	Discuss	In our experiments, we used the Hidden Markov Model (HMM) tagging method described in [Cutting et aL, 1992].

Chapter 3

Citations' Functions Classifier

This chapter present the analysis of the existing machine learning methods to classify citations' context into citations' functions. We analyzed the machine learning classifiers and compared their strengths and weaknesses. Then, in order to select the best machine learning method to improve the citations' functions extraction system.

3.1 Machine Learning Techniques

Athar proposed sentiment analysis of scientific citation [4]. Athar tackled the issue as binary classification. He classified the citations as positive or negative. Since there is no citation corpus to analyze the citation function. ACL Anthology Corpus citation sentences were extracted by the author [14] and 8736 citations from 310 publications were manually labeled. Different features of pre-built corpus are extracted. These features are parts of speech, dependency relations, n-grams scientific lexicon. As main classification methods, Support Vector Machine and Naïve Bayes were used and optimal results were reported using SVM compared to NB. They reported that features of contextual polarity were not working well on the contexts of citations. Good results were achieved by adding negation and dependency relations features. The researcher utilized macro-F metrics for the evaluation of the performance of citations context classification. The approach successfully achieved 0.764 macro-F respectively. It produced the best results by using these features. However, time consuming for feature extraction is the main drawback of the proposed approach and use of implicit citation. The process of manual building of features is too complex and costly.

For addressing these issues, Athar and Teufel [13] constructed a corpus which contains explicit citation sentences and implicit citation sentences. Authors believed that the words and sentences surrounding the citation position contained valuable opinions that could improve the results of detection of the author's purpose in citing works. 203,803 sentences were annotated in four classes. These classes are positive, negative, objective sentences and excluding sentences. Recognizing the implicit citation, good results were presented compared to the use of only explicit citation sentences. While they have tested their methods on acceptable size corpus by using SVM classifier. Their work becomes a domain dependent because they have focused on computational linguistic papers.

The research study of Abu-Jbara et al. [12] employed supervised sequence labeling technique to determine the citation context of a reference and related adjacent sentences which classify citations into six functions; Criticism, Comparison, Use, Substantiating, Basis and Other. The style of references in the journals is different which can affect feature extraction. So that's why, to clean the context of a citation a regular expression was employed. They have extracted four sentences of citations as a window size and annotated into six citations' functions with the help of graduate students. They have employed SVM, Logistic Regression and NB classifiers with the following features; verb, reference count, adverb, self-citation, adjective, dependency relations and negation. The outcome of study revealed that the SVM classifier achieved good results with 0.58 macro-F. However, most of the suggested features specific to citation compared to other studies [15], which have employed established features like POS tags and n-grams.

Parthasarathy et al. extracted citing sentences by using a sentence parser from the data base of Google scholars and Identify adjectives which can be either positive or negative [16]. They suggested that if there is no adjective in a sentence, so the sentence is either unknown or neutral. They used different ML algorithms namely; j48, NB, Sequential Minimal Optimization, AdaBoostM1 to detect the sentiment of citation. A count of research article is 10 paper which are the main inconvenience of their work. The classification algorithms is best suited for the training and the testing of the classifier with large data sets. They also suggested only one feature (obtained adjectives) with supervised learning techniques to detect the sentiments of author.

Sula and Miller have developed a tool for recognizing the sentences of citation and to identify the sentiment of the research article in humanities domain [17]. For the classification of citation they have employed NB algorithm with n-grams model as features. To extract the sentences of citation, four humanities journals were used. They have annotated a few sentences into two different classes, positive and negative then train the NB classifier to categorize the polarity of citation. Unlike previous works, they have extracted context of citation from the domain of humanity research papers, which concentrated on extracting citations sentence from ACL Anthology.

Kim and Thoma suggested a method in the biomedical text documents to detect the sentiment of the author [15]. Their approach includes: (1) extracting the sentences of citation from a research body citing paper and (2) classification of the papers as cited or citing Papers. The kernel function of SVM was applied to classify the polarity of citation in case of n grams and lexicons as a feature set. On 414 titles of biomedical journals SVM was evaluated with a kernel. They categorized the sentiment of the author as positive or others and the performance of their model was 0.90. They have showed better results by using supervised ML approach. However, they concentrated on obtaining explicit citations and neglected valuable citations such as in sections of results and discussion. In addition, their research suffers from the citation context manual annotation process.

Xu et al. concentrated on 285 clinical trial papers in the discussion section and established a rule-based method for citation extraction [18]. In addition, three annotators manually annotated more than 4000 citations. Using SVM supervised machine learning algorithm, they used n-grams and sentiment lexicon features to categorize the citation sentences into two classes positive or negative. Combining their features, the macro F was better than utilizing individual features. The best achievement of this study is their approach of annotating citation but the drawback of this citation annotation process is that it performed manually.

Butt et al. proposed a window-size of five sentences method for extracting the sentences of citation [11]. The authors have employed NB classifier to classify the sentences of citation into two classes, positive and negative sentiments. For data set construction, they used manual annotation process. Their approach obtained 0.80 accuracy which is good. A large window size has been utilized for citation context length and they believe that such size is ideal for conveying the sentiments of authors. In addition, they have utilized only syntactic features and did not attempt to utilize the semantic feature which is helpful for recognizing the sentiments of author.

Hernandez-Alvarez and Gomez have recently proposed a new annotation methodology to label sentences of citation into six classes such as; based on, useful, acknowledge, contrast, weakness and hedges in order to address the classification of citations' functions [19]. They have employed keywords and semantic patterns are semantic features to distinguish the citations' functions. They have developed their own corpus, a corpus for experiments manually and then used for classification of citations. SVM was tested on already annotated corpus and have achieved 0.870 of F1. They utilized explicit and implicit citation in order to identify the functions of citations, but their work is only belong from single domain.

The authors tackled the problem of polarity classification by using information about a reputation of an author [20]. They suggested using various features are unigram, author's id, polarity distribution and p-index. The best performance was reported by combining authors ID and p-index. This study was the first one to improve conventional methods of citation analysis for evaluating the quality of research. However, their research still requires technical skills to use better features to detect the sentiments of scientific citation. We have done the critical analysis of ML techniques which is presented in Table 3.1. In this table, there are six columns first tells the name of scheme, second data sources and so on.

Table 3.1: Critical Analysis of Existing Approaches

Scheme	**Data Sources**	**Classifiers**	**Results**	**Strengths**	**Weaknesses**
Athar [4] 2011	ACL Anthology	SVM	Macro-F: 0.764	Best results with combining ngrams and dependency relations.	Time consuming for feature extraction, did not handle implicit citation, citation annotation (manually), did not compare results with other approaches. Low results.
Athar [13] 2012	ACL Anthology	SVM	Macro-F: 0.731	Using citation context length (explicit and implicit), Improved results with different context windows, Best results with combining ngrams and dependency relations.	Time consuming for feature extraction, citation annotation (manually), only Focused on computational linguistic papers.
Kim and Thoma [15] 2015	Own	SVM	Macro F: 0.90	Best results with combining uni-gram and bi-gram.	Did not use valuable citations such as in sections of results and discussion. Did not handle negation problem.

Table 3.1 - Continued from Previous Page

Scheme	**Data Sources**	**Classifiers**	**Results**	**Strengths**	**Weaknesses**
Xu et al. [18] 2015	Google scholar	SVM	Macro-F: 0.719	Combination of n-grams and sentiment lexicons features to achieve better results.	Annotation process is conducted manually. Citation analysis only for biomedical publications
Hernandez-Alvarez and Gomez [19] 2015	ACL Anthology	SVM	Macro-F: 0.870	keywords and semantic patterns	citation annotation (manually)
Ma et al. [20] 2016	Athar and Teufel (2012)	SVM	Macro-F: 0.645	To improve H-index method by including negative polarity in the calculation process.	Did not handle implicit citation. Citation annotation (manually). Did not compare results with other approaches.
Abu-Jbara et al. [12] 2017	ACL Anthology	SVM	Macro-F: 0.58	Best results with dependency relations. reference count and closest verb, adjective and adverb to the target reference .	Only Focused on computational linguistic papers. Citation annotation (manually). Low results.

3.2 Selection of Machine Learning Methods

Machine learning methods have been used extensively for citations analysis. As it is evident from the literature review that most of the approaches utilized Naive Bayes (NB) and Support Vector Machine (SVM) to classify the citation texts into Citations Functions. Most of the approaches that address the citations' functions issues are applying different classifier, which is trained on various features of citation context. From the literature we have selected these two classifiers which are widely used; SVM and NB. These classifiers performed very well on text classification. In this study, we have carried out experiments using these widely used machine learning classifiers. Along with these classifiers, we also adopted Random Forest (RF) for citations' functions classification. As we were advised in proposal defense to use this classifier because it was perceived that RF perform well on text classification.

Chapter 4

CiFE - Citation Function Extraction from Citations' Contexts

As it is clear from the literature review that most of the approaches to extract citations functions classify these functions into three macro-level categories such as: 1) positive, 2) negative and 3) neutral. These functions can be divided into multiple sub functions as is done in CCRO [6]. However, extraction of these Functions from the literature is a challenge which is being addressed in this thesis. In this chapter, we present a detailed methodology proposed to develop CiFE- Citation Function Extraction from Citations' Context to extract the citations functions proposed in CCRO from the research papers. Each step of CiFE is described in the following sections. This whole process have been completed in four steps. first of all, we selected Athar's data set and manually annoted it into eight CCRO classes. Then at the second step, feature selection rules were devised for extraction of important features from citation texts. At third step, machine learning classifier was trained on annotated data set. At the fourth and final step, a trained machine learning classifier was used for the prediction of citation text. In this way, this complex process was completed and the Figure 4.1 is a graphical representation of our proposed System.

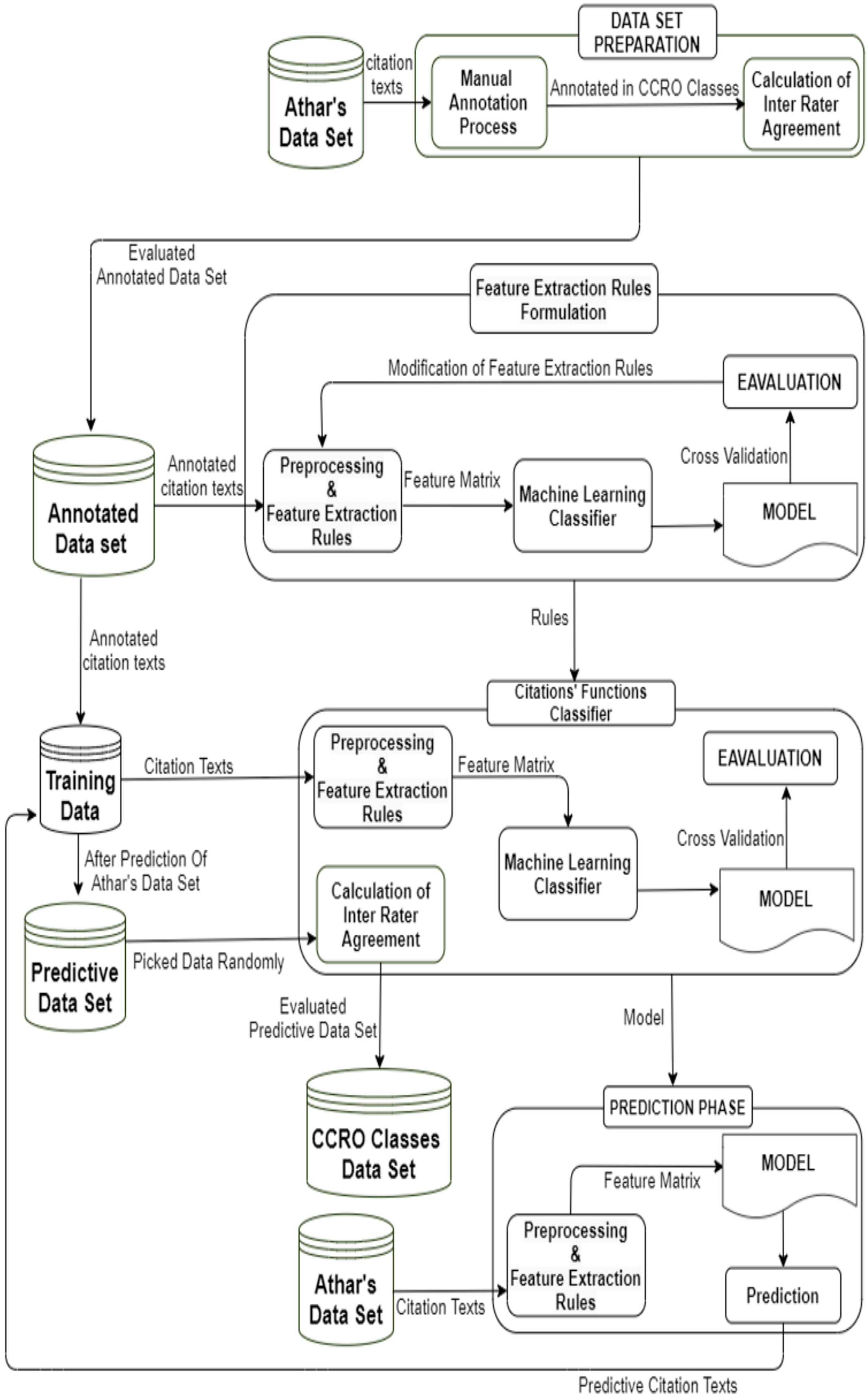
DATA SET PREPARATION
Athar's Data Set
citation texts
Manual Annotation Process
Annotated in CCRO Classes
Calculation of Inter Rater Agreement
Evaluated Annotated Data Set
Feature Extraction Rules Formulation
Modification of Feature Extraction Rules
EAVALUATION
Annotated Data set
Annotated citation texts
Preprocessing & Feature Extraction Rules
Feature Matrix
Machine Learning Classifier
Cross Validation
MODEL
Rules
Annotated citation texts
Citations' Functions Classifier
Training Data
Citation Texts
Preprocessing & Feature Extraction Rules
Feature Matrix
EAVALUATION
Machine Learning Classifier
Cross Validation
MODEL
After Prediction Of Athar's Data Set
Calculation of Inter Rater Agreement
Predictive Data Set
Picked Data Randomly
Evaluated Predictive Data Set
Model
CCRO Classes Data Set
PREDICTION PHASE
MODEL
Feature Matrix
Preprocessing & Feature Extraction Rules
Athar's Data Set
Citation Texts
Prediction
Predictive Citation Texts

Figure 4.1: Diagram of Proposed System

4.1 Data Set Preparation

We have used supervised machine learning system which requires annotated data set. Annotated data set helps to train the ML classifier in classifying the citations' texts into CCRO classes. For annotating citation context manual based approach can be used [12]. We randomly selected a set of three hundred sentences from Athar's data set [4]. Which were annotated in three macro level classes positive, negative and neutral but we have to annotate these sentences in to eight micro level classes by annotators. By doing this, we can train machine learning classifier to automatically annotate the new citation texts into CCRO classes. All steps of data set preparation are presented in Figure 4.2.

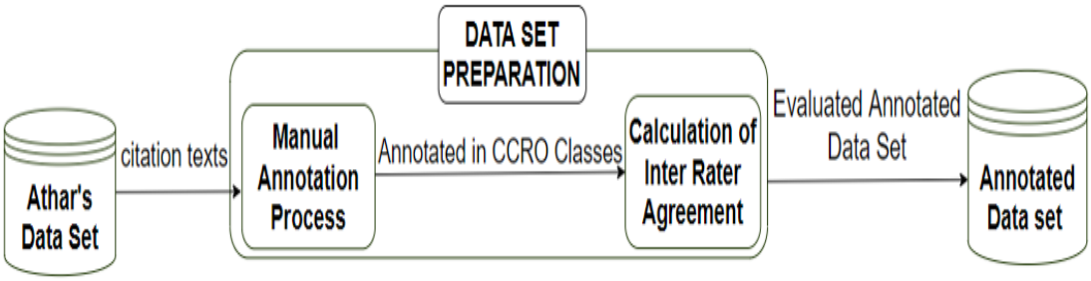
DATA SET PREPARATION
Athar's Data Set
citation texts
Manual Annotation Process
Annotated in CCRO Classes
Calculation of Inter Rater Agreement
Evaluated Annotated Data Set
Annotated Data set

Figure 4.2: Data Set Preparation

The selection of data set is a very important step in a complete evaluation of our proposed technique. To evaluate the proposed technique, a diversified data set is required to conduct the research. ACL Anthology Corpus is one of the important citation sources which are commonly used to analyze the scientific citations' classes [14][21]. In our experiments, we have employed a specific version of the AAN data set. The ACL Anthology Network (AAN) is a huge and manually compiled corpus which covers all documents published over four decades by two journal such as ACL and computational linguistics. Some researchers have used this corpus to construct their own data sets. One of another corpus has been designed and annotated by Athar [4] which contain 8,736 AAN citation sentences labeled as Citing Paper ID, Cited Paper ID, Citation Text, and three sentiment classes which are positive, negative and neutral. There are 830 positive citations, 281 negative and 7,625 objective citations. This is consistent with earlier work [22], which shows that citations are mostly neutral sentimentally. Table 4.1 provides a highly skewed distribution of the sentiment classes in the corpus, 87.3% of which is objective and just around 12.7% having some sentiments.

TABLE 4.1: Distribution of Athar's Data Set

Sentiment classes	**Count**	**percentage**
Positive	830	9.5%
Negative	281	3.2%
Neutral	7625	87.3%

After the selection of citation texts from Athar's Data Set [4], we manually annotated it in CCRO classes. In order to annotate the data set, first of all, we developed an instruction table to facilitate the annotators in the process of annotation (Table 4.2). Afterwards, to assist the annotators to label the citation text in to 8 reasoning classes, we gave 5, 5 examples for each citation reason class. The table have four attributes consisting on context class, sub class, description and examples. First attribute of this table was context class which defines the polarity of citation text. The second attribute defines the eight classes of citation text. The third column further defines the CCRO classes. The fourth column consists of examples. We assigned this task to two groups of students, each group consisted of 2, 2 students graduated in English linguistics. They were given 300 randomly selected sentences from Athar's data set and were asked to label these sentences into CCRO classes. They were also provided a table of instructions with explanations and examples. Further, they were asked to underline the words and phrases which helped them to annotate the sentence in their relevant class.

To check and evaluate the accuracy of Annotation process between two groups of annotators, we have utilized the Cohen Kappa Coefficient to evaluate the agreement [23]. The inter-annotator agreement kappa coefficient is the primary tool for measuring the consistency of the annotation process. In order to interpret the values of kappa we used a scale proposed by Landis and Koch [24]. This scale is shown in Table 4.3.

Table 4.2: Annotation Scheme for Citations' Functions

Context Class	**Citations' Functions**	**Rules**	**Examples**
Positive	Incorporate	When citing sentence is annotated as Incorporate its mean citing paper take in or contain something to make part of a whole of the cited paper.	Smith(2007) **describe** a **more efficient algorithm** that can compute all edge expectations in O(n3) time using the inverse of the Kirchoff matrix K1.
	Extend	When citing sentence is annotated as Extend its mean citing paper extend the knowledge of the cited paper.	Stochastic models (Cutting et al., 1992; Dermatas et al., 1995; Brants, 2000) have been **widely used** in POS tagging for simplicity and language independence of the models.
	Based On	When citing sentence is annotated as Based On its mean citing paper to make a decision by using particular ideas or facts of the cited paper.	One of the most effective taggers **based on** a pure HMM is that **developed at** Xerox (Cutting et al. , 1992).
Negative	Negate	When citing sentence is annotated as Negate its mean citing paper consider the work as ineffective or invalid of the cited paper.	Therefore, sublanguage techniques such as Sager(1981) and Smadja (1993) **do not work.**
	Criticize	When citing sentence is annotated as Criticize it means citing paper finds the fault and then point out the fault of the cited paper.	Chiang(2005) introduced a constituent feature to reward phrases that match asyntactic tree **but did not** yield significant improvement.

Table 4.2 - Continued from Previous Page

Context Class	**Citations' Functions**	**Rules**	**Examples**
	Contrast	When citing sentence is annotated as Contrast it means citing paper shows differences with opposite nature between the work of citing and cited paper.	With all but two formats IBI-IG achieves FZ=l rates **better than** the best published result in (Ramshaw and Marcus, 1995).
Neutral	Compare	When citing sentence is annotated as Compare its citing paper showing a comparison between the work of citing and cited paper.	Actually, it is defined **similarly to** the translation model in SMT (Koehn et al., 2003).
	Discuss	when citing sentence is annotated as Discuss if it is a neutral description of the cited paper or if it does not fall under any of the categories mentioned above.	In our experiments, we **used** the Hidden Markov Model (HMM) tagging method **described** in [Cutting et al, 1992].

Table 4.3: Interpreting Kappa Values

Kappa value	**Agreement level**
0.00	Poor
0.01-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

4.2 Pre-Processing

Pre-processing is a technique which remove unnecessary and noisy data from the data set. Data sets are generally incomplete: containing noisy data (meaningless data), lack of attribute values (Missing Values) etc. We implemented different steps for pre-processing for example Tokenization, Noise Removal, Stop Word's Removal, Lemmatization and POS Tagging. All these steps have been discussed one by one:

4.2.1 Tokenization

The first pre-processing step is Tokenization. In this step the citation texts can be divided into meaningful pieces. These pieces are known as tokens. For example, we can split a chunk of text into words, or we can split it into sentences. We have split the citation texts into words. We use the Spacy for tokenization, which is the best-known and most widely used Natural Language Processing library [25].

4.2.2 Noise Removal

It is important to remove noise from data because it can adversely affect accuracy. The data sets generally contain noise such as unnecessary punctuation and null values. Different methods exist for removing noise such as manually filling the missing values, filling using calculated values and ignoring the missing records. However, we ignored these citation texts because it is the simplest and most efficient way to handle the missing data. After tokenization, some punctuations

4.2.3 Stop Words' Removal

Stop words in a language are the words which are most frequently occur such as on, of, a etc. These words have no significant meaning, so they must be removed from the citation texts for correct measurement. We used Spacy library to remove stop words from all the citations texts, because this library contains a stop words list. Spacy library compares the tokenized list to its own list of stop words then stop word removal from the corpus has been performed. This step is useful in reducing the dimension of a features space.

4.2.4 Lemmatization

Lemmatization is a way to reduce words to their roots or basic words. The benefit of lemmatization is that it decreases the size of vocabulary. For example, all the terms like program, programs, programmer, programing, and programmers are lemmatized into their root word program. We have done this using Spacy library that transforms each word of citation texts to its root words. For all citation texts the lemmatization algorithm is applied.

4.2.5 POS Tagging

Parts of Speech (POS) tagging is basically utilized for to eliminate uncertainty by clarifying something. Moreover, it is the process of defining a word to a particular part of speech in a text. The main usage is for the purpose of selecting linguistic features. We can find the Linguistic features very easily with POS tags. These features include adjective, verb, adverb and their distinctive types. Such linguistic features are commonly seen as indicators of sentiments. Spacy library is used to apply on the citation texts in which help us to define a word to a particular POS in a text. We extracted important linguistic features from POS tagged words, which are further used for classification of citation texts.

4.3 Feature Extraction Techniques

In the classification of citation texts, the extraction of features is an important technique. The functions of citations can be identified by using these techniques.

In the previous chapter we have discussed several types of features for citation analysis. Here, we describe the most common features which have been used for the classification of citation texts.

4.3.1 N-grams

An n-gram is a set of contiguous terms in a given text. The character 'n' refers to the sequence length. When n = 1, the series is considered a unigram, so if n = 2 or 3, it is considered a bigram or trigram. The n-grams are as follows for the simple sentence Ali is a good student:
unigrams: Ali, is, a, good, student
bigrams: Ali is, is a, a good, good student
trigrams: Ali is a, is a good, a good student
4-grams: Ali is a good, is a good student
In existing sentiment classification tasks for movie reviews, length 1 and length 2 of N-grams performed extraordinary [26]. Bigrams with adjectives and adverbs are considered to be more sentimental. In addition to looking specifically for the scope of the negation words[27]. We have used n-gram lengths from 1 to 2 for experiments.

4.3.2 Bi-Tagged

Bi-tagged type features are obtained by POS tagging. The information based on POS is utilized for extracting sentiment-rich features, although adjectives and adverbs have been investigated in literature, the nature of these are subjective. One of the researchers Turney proposed a technique for extracting Bi-word sentiment-rich features in such a way that its one member is either belong to adjective or adverb, for example, adjective-noun, adverb-adjective, noun-adjective, adverb-verb [28]. We have also observed that the verb (verb-noun, verb-adjective, adjective-verb, and adverb-verb) can also provide reasoning information that is useful for citations' functions classification.

4.3.3 Dependency Features

Dependency features describe the grammatical relation between the words. Each feature in the dependency structure represents a binary relationship between a head word and a dependent word. Generally dependencies described as triples form relation (head, dependent). As it is illustrated in the following sentence and also presented in Figure 4.3:
Our system outperforms competing approaches.
This above sentence contains 4 tokens corresponding to the following triplets.

1. poss (system, our)
2. nsubj (outperforms, system)
3. amod (approaches, competing)
4. dobj (outperforms, approaches)

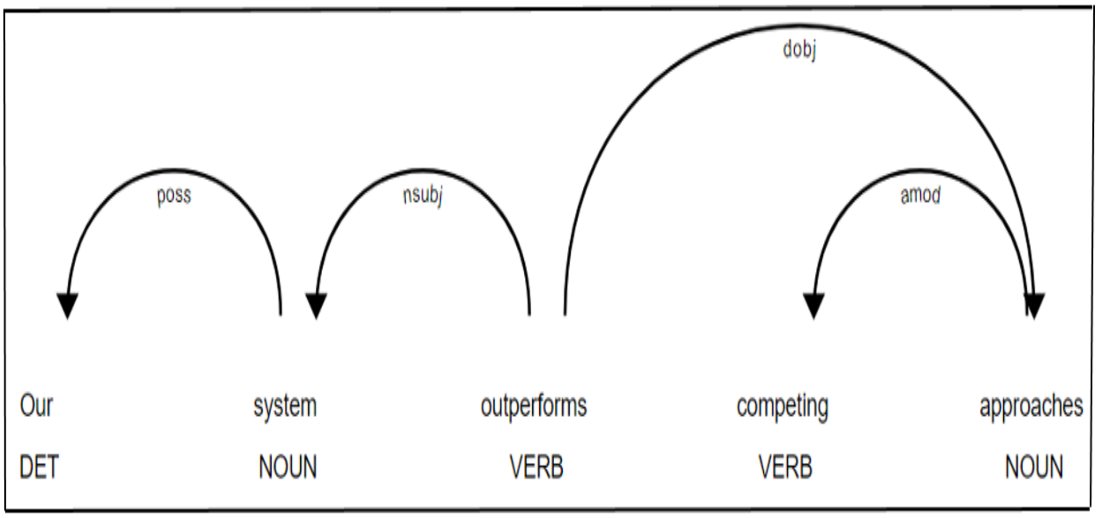

FIGURE 4.3: Dependency Structure

Showing dependency relationships are very helpful for citation analysis and numerous scholars have focused on utilizing nsubj, advmod (adverb modifier) and amod (adjectival modifier) information in their systems [4][27][29][30]. These tags are also indicators of subjectivity in a sentence. The above researchers motivate us to use the dependency structures capturing the long-distance relationship between words.

4.3.4 Window Based Negation

Negations are very important in linguistics because they have an effect on the polarities of other words. Negations contain terms like no, not, shouldn't, etc. Moreover, in case of negation in a sentence, it becomes necessary to identify words sequence affected by this term. There has been a lot of work which deals with negation and its scope in the Citations' functions classification. We used a negation list which contained 31 terms (no, not, rather, couldn't, wasn't, didn't, wouldn't, shouldn't, weren't, don't, doesn't, haven't, hasn't, won't, wont, hadn't, never, none, nobody, nothing, neither, nor, nowhere, isn't, can't, cannot, mustn't, mightn't, shan't, without, needn't) [27]. We have detected the negation terms by using negation list and dependency tree. For the scope of negation, we have followed the window-based approach [4] [26]. We have used a negation window of 7 words for citations' functions classification. All words within a 7-word range of any terms of negation are suffixed with a token-neg to differentiate between them and other versions.

4.4 Feature Selection Rules

We devised feature selection rules for extraction of important features from citation texts to improve the accuracy of citations' functions classification. These rules have been prepared by conducting two experiments on annotated data set. Firstly, during the annotation process, the annotators were asked to underline the words and phrases which helped them to annotate the citation texts in their relevant classes. Afterwards, we found that most helpful part of speech for the annotators were verbs, adverbs, adjectives. Thus, these POS helped to figure out the reasoning class of citations. Secondly, we identified some important parts of speech from literature review which performed well in text classification. These parts of speech include nouns, verbs, adverbs and adjectives. We have performed experiments on annotated data set with these important parts of speech. After analyzing the results, we found that there are some parts of speech which were more beneficial than other parts of speech. In this data set, these parts of speech include adjectives, verbs and adverbs which informs well about the relevant citations' functions. In this way, we devised feature selection rules for the extraction of important features from citation texts. These feature selection rules are given in Table 4.4.

TABLE 4.4: Features Extraction Rules

Sr #	**Rules**	**Examples**	**Features**
1	*If a verb holds nominal subject (nsubj) dependency with a subject, then pick the verb in a citation text.*	He has achieved state-of-the-art results by applying M.E. to parsing (Ratnaparkhi, 1997a) and part-of-speech tagging (Ratnaparkhi, 1996).	VERB: achieve
2	*If a word holds adverbial modifier (advmod) dependency with a word then pick a bigram and unigram.*	Smith and Smith (2007) describe a more efficient algorithm that can compute all edge expectations in O(n3)time using the inverse of the Kirchoff matrix K1.	advmod: more_efficient, ADV: more, ADJ: efficient
3	*If a verb occurs before or after the citation anchor with in window size four then pick the verb.*	He has achieved state-of-the art results by applying M.E. to parsing (Ratnaparkhi, 1997a) and POS tagging (Ratnaparkhi, 1996).	VERB: parse, VERB: tag

Table 4.4 - Continued from Previous Page

Sr #	Rules	Examples	Features
4	*If an adjective (ADJ) occurs immediately after verb then pick the ADJ.*	There has been significant work with such models for greedy sequence modeling in NLP (Ratnaparkhi, 1996; Borthwick et al. , 1998).	ADJ: significant
5	Pick seven words after negation clause or contraction clue before the punctuation mark. Do not consider stop words.	As the tagger of Ratnaparkhi (1996) cannot tag a word lattice, we cannot back off to this tagging.	Negated words: cannot, not tag, not word, not lattice, not back, not off, not tag
6	*6.1: Pick conjunction (conj) label words.* *6.2: If adposition (ADP) or subordinating conjunction (sconj) label words occur at the start of the sentence then pick it.* *6.3: If adposition (ADP) label words occur after the punctuation mark then pick the ADP.* *6.4: If adverb (ADV) occurs before the punctuation mark then pick the ADV.*	Hanks and Church (1990) proposed using point wise mutual information to identify collocations in lexicography; however, the method may result in unacceptable collocations for low-count pairs.	ADV: however, CONJ: and
7	*If a word holds nominal subject (nsubj) dependency after adjective (ADJ) or adverb (ADV), then pick ADJ or ADV.*	An alternative method (Wu, 1997) makes decisions at the end but has a high computational requirement.	ADJ: alternative

Table 4.4 - Continued from Previous Page

Sr #	Rules	Examples	Features
8	If adjective (Adj) or adverb (Adv) occurs before subordinating conjunction (SCONJ) or adposition (ADP) then pick its bigram ADJ_ADP or ADV_ADP and pick ADJ or ADV.	The model we use is similar to that of (Ratnaparkhi, 1996).	ADJ_ADP: similar_to ADJ: similar
9	If adverb (ADV) or adjective (ADJ) occurs immediately after nominal subject (nsubj) dependency verb then pick ADV or ADJ.	There are however other similarity metrics (e.g. BLEU (Papineni et al., 2002)) which could be used equally well.	ADV: equally
10	If a verb occurs immediately after or before the adposition (ADP) or preposition (prep) then pick both verb and ADP or Prep as a bigram.	Ramshaw and Marcus (1995) approached chunking by using a machine learning method.	ADP_VERB: by_use VERB_ADP: chunk_by

4.5 Vectorization

Most of the machine learning algorithms often take numeric vector as an input. However, before performing any operation on a text, we need a way to convert each citation text into a numeric vectors. This is one of the fundamental problem in data mining, which aims to numerically represent the unstructured text documents to make them mathematically computable. For this purpose, we used feature matrix which used for text vectorization. In feature matrix, each word is converted into a binary value 1 or 0, which indicate the word appear in a citation text or not. Several types of features that capture the characteristics of citation sentences are extracted by devised feature extraction rules are served as the inputs of automatic classifiers. Let us consider an example to understand the working of feature matrix. There are two citation texts that contain terms such as:

1. Sublanguage techniques do not work.

2. The model we use is similar to (Ratnaparkhi, 1996).

First we have to form dictionary of unique words by using the devised feature selection rules from these citation texts such as: (VERB: work, donot, neg:work, VERB:use, ADJ:similar, similar_to). Then to make the vector of the first citation text, the terms of the citation text are matched with dictionary words. If term matched placed '1' in that index if not then placed '0'. Example represented in the Table 4.5.

Table 4.5: Vectorization

Features	**VERB: work**	**donot**	**neg:work**	**VERB: use**	**ADJ: similar**	**similar_to**
Citation Text 1 Vec	1	1	1	0	0	0
Citation Text 2 Vec	0	0	0	1	1	1

4.6 Classification Techniques

Mostly approaches presented in literature review addresses the Citations' Functions issues by applying ML classifier which trained on citation context features. We experimented by using these widely used ML classifiers such as: 1) Naive Bayes (NB), 2) Random Forest and 3) Support Vector Machine.

4.6.1 Naïve Bayes (NB)

The NB classifier [31] works on the basis probability calculation of data. It assumes that the existence of a specific feature in a class is irrelevant to the existence of any other feature. It shows better performance with multi-class problems as well as perform better in text classification. Moreover, the model is simple to create, and particularly useful for very huge data sets.

4.6.2 Random Forest (RF)

The RF classifier [32] is a learning based algorithm which is commonly used for classification on labeled data sets. A forest is composed of trees and more trees means stronger forest. Similarly, RF makes a decision tree on each data set, after that each of them gets prediction and finally selects the best solution by voting method. This approach is ensemble based which is better than a single decision tree approach, and by averaging the result it reduces the over fitting.

4.6.3 Support Vector Machine (SVM)

Similar to RF, the SVM [33] are one of the powerful and flexible learning based algorithm which is commonly used for classification task on labeled data sets. The representation of classes in SVM model is on hyperplane in multidimensional space. The error chances in hyperplane is very low because it is generated in an iterative manner by SVM. SVM's main objective is to classify the data sets into classes. For finding maximum marginal hyperplane which can be achieved in two steps. The first one is, SVM iteratively generates hyperplanes which segregate classes in a best way. Then, it chooses the hyperplane that correctly separates the classes. SVM classifiers have excellent precision and function well with high dimensional space. Basically, SVM classifiers use subset of training points thus uses very less memory.

4.7 Evaluation

In the proposed technique, the main objective of evaluation is to identify the impact of linguistic features as well as ML classifiers for detecting citations' functions. For experimental purpose we have used Weka tool for classification [34]. In ML classifiers we have chosen three different machine learning classifiers such as: 1) SVM, 2) RF and 3) NB. For experiments we have used the training/testing data set in a 10-fold cross validation mode. To calculate the results of our proposed technique, the standard formula of precision, weighted-average precision, macro-precision, recall, weighted-average recall, macro-recall, F1-score, weighted-average F1-score and macro F is calculated. The evaluation parameter have used for the multi class classification are given below:

$$Precision(P) = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{4.1}$$

$$Weighted\ Avg(P) = \frac{1}{total\ samples}\sum_{i=1}^{n}((samples\ of\ class\ i) * Pi) \tag{4.2}$$

$$Macro - Precision = \frac{1}{n}\sum_{i=1}^{n} Pi \tag{4.3}$$

$$Recall(R) = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{4.4}$$

$$Weighted\ Avg(R) = \frac{1}{total\ samples}\sum_{i=1}^{n}((samples\ of\ class\ i) * Ri) \tag{4.5}$$

$$Macro - Recall = \frac{1}{n}\sum_{i=1}^{n} Ri \tag{4.6}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4.7}$$

$$Weighted\ Avg(F1 - score) = 2 * \frac{Weighted\ Avg(P) * Weighted\ Avg(R)}{Weighted\ Avg(P) + Weighted\ Avg(R)} \tag{4.8}$$

$$Macro\ F = 2 * \frac{Macro\ Avg(p) * Macro\ Avg(R)}{Macro\ Avg(P) + Macro\ Avg(R)} \tag{4.9}$$

Chapter 5

Results and Evaluation

In the previous chapter, we have explained the in-depth details of the proposed classification system. This chapter presents the details about the results that have been obtained by applying the proposed classification system.

5.1 Evaluation of Annotated Data Set

In our experiments, we used a specific version of the AAN data set. We have discussed about this data set in detail in the previous chapter. Some researchers have used this data set to construct their own data sets. One of another corpus has designed and annotated by Athar [4] which contain 8,736 AAN citation sentences labeled as Citing Paper ID, Cited Paper ID, Citation Text, and three sentiment classes which are positive, negative and neutral. There are 830 positive citations, 281 negative and 7,625 objective citations. Our data set is based on Athar's data set. We have randomly selected a set of three hundred sentences from this data set. The distribution of randomly selected data contains 100 positive, 100 negative and 100 neutral sentences. We have assigned manual annotation task to two groups of students. Each group consisted of 2, 2 students graduated in English linguistics. They were given these selected sentences and were asked to label these sentences into eight citations' functions. To help the annotators, we have assigned them a table of instruction with explanations and examples. We need to calculate inter-annotator agreement between two groups of annotators as shown in Table 5.1.

Table 5.1: Contingency Table for Calculating Kappa

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**Total**
A	29	4	3	0	0	0	0	0	**36**
B	4	28	0	0	0	0	0	0	**32**
C	2	1	22	0	0	0	0	0	**25**
D	0	0	0	18	4	3	0	0	**25**
E	0	0	0	4	41	2	0	0	**47**
F	0	0	0	1	5	33	0	0	**39**
G	0	0	0	0	0	0	35	4	**39**
H	0	0	0	0	0	0	2	58	**60**
Total	**35**	**33**	**25**	**23**	**50**	**38**	**37**	**62**	**303**

For evaluation of inter annotator agreement we have used the Cohen Kappa Coefficient to evaluate the agreement. We discussed in previous chapter about kappa calculation process. We calculated Kappa: 0.851, on Landis and Kochs [24] scale, the Kappa value indicates almost perfect agreement.

5.2 Evaluation of Preprocessing

After completion of annotation process, all of these citations needed to be cleaned. The following steps have been carried out while pre-processing:

1. Tokenize the text of all citations on the basis of space. The tokenization has been performed by using Spacy Library.
2. Removal of stop words from all the citations using Spacy stop word list.
3. Conversion of all the text of citations into their root terms by using Spacy Lemmatization algorithm.

Afterwards, 30% of annotated data set were manually verified. In this way, we have performed all the pre-processing steps successfully.

5.3 Evaluation of Important Features Extraction

We have conducted a number of experiments after the completion of annotation process to identify the important features for accurate classification. The experiment was consisted in two rounds: The first round is Features Extraction without using Polarity and Rules and the second round is Features Extraction with using Polarity and Rules. The first experiment was done in order to identify the important features. We have devised rules with the help of these important features. Along with this, the underlined words and phrases during annotation process also helped in devising rules. During the second experiment, we have analyzed the results of rules and then modified these rules. We have observed that, the results were improved after modification. The Figure 5.1 is a visual representation of these experiments.

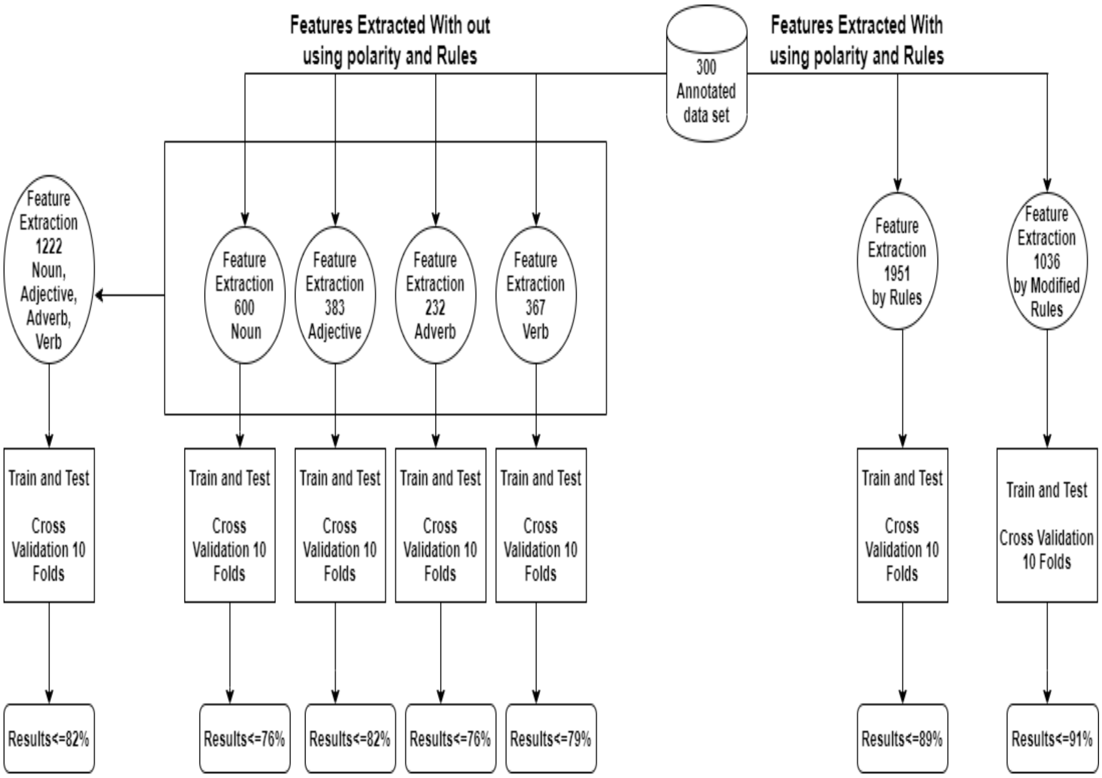
Features Extracted With out using polarity and Rules
300 Annotated data set
Features Extracted With using polarity and Rules
Feature Extraction 1222 Noun, Adjective, Adverb, Verb
Feature Extraction 600 Noun
Feature Extraction 383 Adjective
Feature Extraction 232 Adverb
Feature Extraction 367 Verb
Feature Extraction 1951 by Rules
Feature Extraction 1036 by Modified Rules
Train and Test Cross Validation 10 Folds
Train and Test Cross Validation 10 Folds
Train and Test Cross Validation 10 Folds
Train and Test Cross Validation 10 Folds
Train and Test Cross Validation 10 Folds
Train and Test Cross Validation 10 Folds
Train and Test Cross Validation 10 Folds
Results<=82%
Results<=76%
Results<=82%
Results<=76%
Results<=79%
Results<=89%
Results<=91%

Figure 5.1: Important Features Identification Architecture

5.3.1 Identification of Important Parts of Speech

There are a number of parts of speech in a citation text. We have identified such parts of speech from literature review, which performed well in text classification. These parts of speech include nouns, verbs, adverbs and adjectives. We have trained NB, RF and SVM classifier on these features. In the experiments, we have trained our classifier one by one on parts of speech separately. First of all, on nouns, secondly on verbs, thirdly on adverbs and at the last on adjective. Afterwards, we trained these classifiers on these parts of speech collectively. We have used 10-fold cross validation to analyze the results of these classifiers. Afterwards, we took three measurements named weighted-average precision, weighted-average recall, and weighted-average F1-score. With the help of these measurements, we analyzed the result accuracy ratio of the parts of speech shown in Figure 5.2. We have achieved maximum 82% weighted-average F1-score. After analyzing the results, we found that there are some parts of speech are helping more than other parts of speech. In this data set, these parts of speech includes adjectives, verbs and adverbs which informs well about the relevant citations' functions. The result of SVM classifier is outperformed other classifiers. Furthermore, this experiment helped us a lot in the process of feature selection rules.

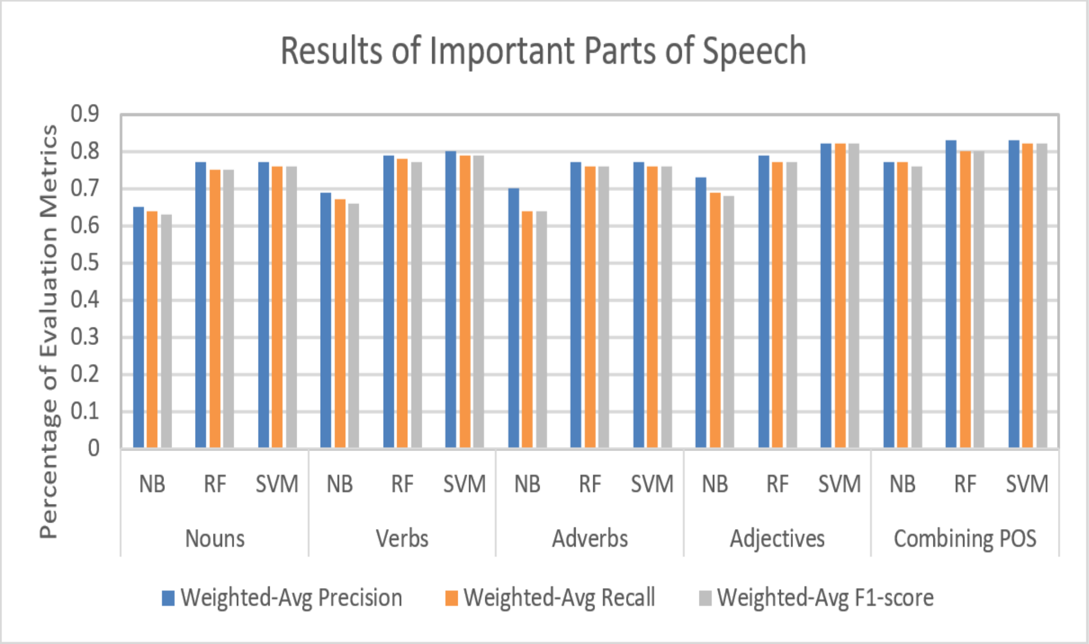

FIGURE 5.2: Results of Important Parts of Speech

5.3.2 Evaluation of Feature Extraction Rules

In the previous chapter, we have discussed feature extraction rules in detail. Several types of features that capture the characteristics of citation sentences are extracted by devised feature extraction rules are served as the inputs of automatic classifiers. We have extracted one thousand and thirty six important features with the help of these rules from the annotated data set. After that, we have trained our classifiers NB, RF and SVM on important extracted features. We used 10 fold cross validation approach to analyze the results of these classifiers. We have employed three standard evaluation measures for the evaluation of the classifiers including weighted-average precision, weighted-average recall, and weighted-average F1-score. With the help of these measures, we analyzed the results of the feature extraction rules shown in Figure 5.3. We have achieved 91% weighted-average F1-score. These results reveals that in classification while using rules have outperformed the results which were not using the rules. Furthermore, the result of SVM classifier has outperformed the results of Naive Bayes(NB) and Random Forest(RF) classifiers.

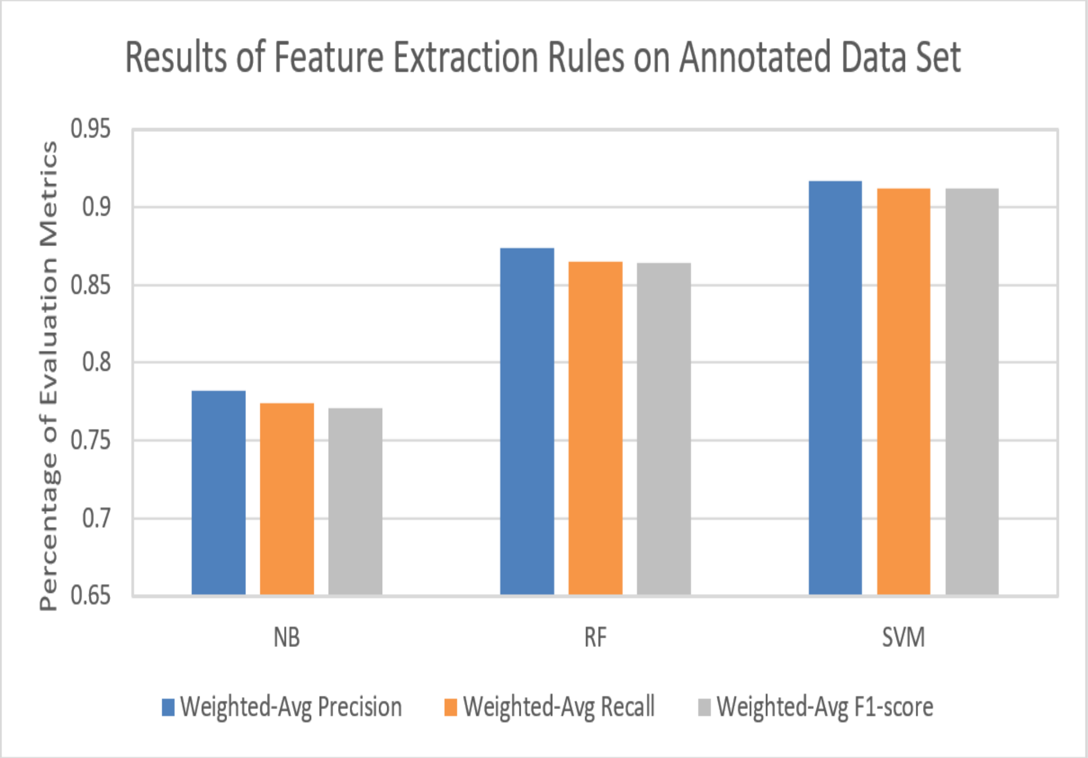

Figure 5.3: Results of Feature Extraction Rules on Annotated Data Set

5.4 Evaluation of Citations' Functions Classification

For the classification of Athar's data set into 8 citations' functions. For this purpose, we have divided this data set into two parts. The first part contains 300 citing sentences while the second part of this data set contains 8428 citing sentences. In the first part, we devised feature extraction rules and selected the SVM classifier for the prediction of the second part of data set. We have discussed in details about the results of feature extraction rules in the previous steps. Along with this, the performance of the classifiers was also evaluated. In the second part, we divided 8428 sentences into 17 sub sets. Each sub set consisted on 495 sentences. We have predicted all citations in 17 steps by using SVM classifier based on the principle of train and test. In this way, we have completed prediction process of positive (724), negative (182) and neutral (7522) sentences of the second part. During the prediction process, we have picked up 10% data randomly at first 5 prediction steps. We have checked the results and found that 90% of the predicted sentences were correct on average. The Figure 5.4 is a graphical representation of this predicted data set into eight Citations' functions.

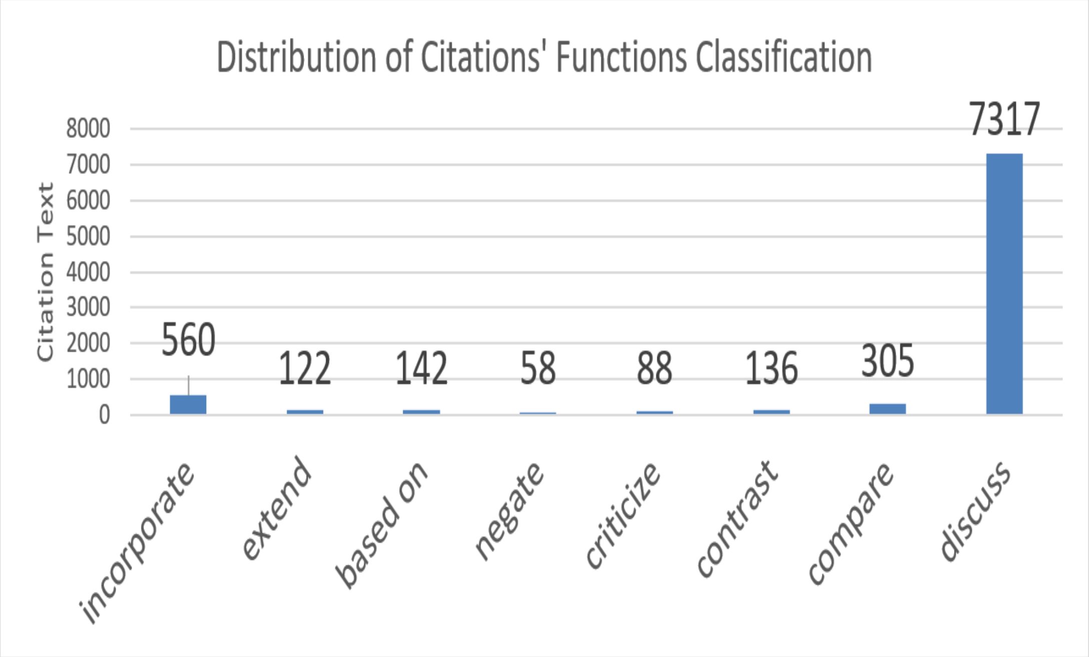

FIGURE 5.4: Distribution of Citations' Functions Classification

After the completion of prediction, we have picked 15% citing sentences randomly from the predicted data. We have passed these sentences to English Linguistic annotator and asked them to annotate these sentences into 8 citations' functions. After manually annotation of 15% citing sentences. We calculated inter annotator agreement between annotator and machine predictions shown in Table 5.2.

Table 5.2: Contingency Table for Citations' Functions Classification

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**Total**
A	92	8	5	0	0	0	1	0	**106**
B	5	13	0	0	0	0	0	0	**18**
C	7	4	11	0	0	0	0	0	**22**
D	0	0	0	8	10	6	0	0	**24**
E	0	0	0	0	24	0	0	0	**24**
F	0	0	0	3	12	56	0	0	**71**
G	0	0	0	0	1	0	35	8	**44**
H	0	0	0	0	0	0	84	921	**1005**
Total	**104**	**25**	**16**	**11**	**47**	**62**	**120**	**929**	**1314**

For evaluation of inter annotator agreement we have used the Cohen Kappa Coefficient to evaluate the agreement. We discussed in previous chapter about kappa calculation process. We calculated Kappa: 0.737, on Landis and Kochs [24] scale, the Kappa value indicates substantial agreement.

Furthermore, we have applied several ML classifiers including SVM, Random Forest and Naïve Bayes for experiments. For experimental setup we have used the training/testing data set in a 10-fold cross validation mode with twelve thousand and two hundred features for the purpose of evaluation of citations' functions classification. We achieved 98% weighted-average F1-score with SVM classifier. Weighted-average precision and weighted-average recall have same value because due to the nature of the data number of false positives is same as the number of false negatives. In this experiment the value of three metrics has same while in other experiments as shown in Figure 5.2 and 5.3, the values vary. The outcome of our experiments revealed that the SVM classifier outperforms the Random Forest and Naive Bayes classifiers. To evaluate the results of our predicted data set, the standard formula of weighted-average precision, weighted-average recall and weighted-average F1-score were calculated, as is shown in Figure 5.5.

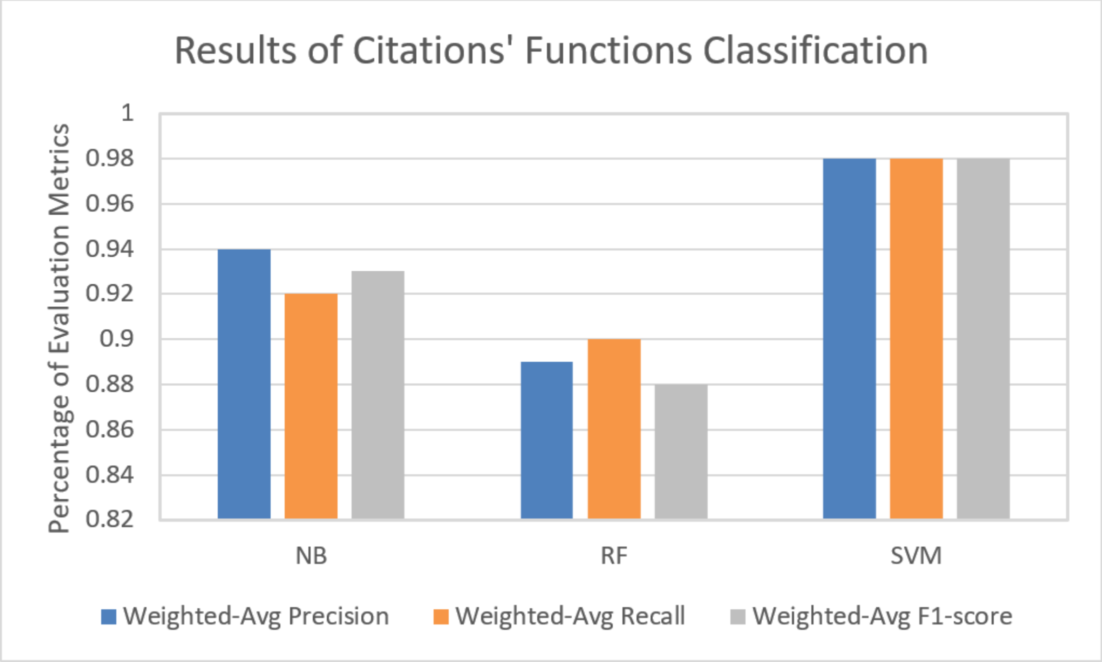

FIGURE 5.5: Results of Citations' Functions Classification

5.5 Comparison with Other System Results

The citation analysis community has proposed multiple approaches for performing citations' functions classification. Most of these approaches have utilized different types of citations' functions, features and machine learning techniques described in literature review chapter. In this thesis we have utilized 8 citations' functions from CCRO ontology. We have used feature selection rules to extract important features from citation texts and best machine learning classifiers NB, RF and SVM to improve the results of citations' functions classification. We have used 10-fold cross validation to analyze the results of these classifiers. Afterwards, we have evaluated our results on the basis of three measures such as weighted-average precision, weighted-average recall and weighted-average F1-score. For comparison purpose, we have used macro precision, macro recall and macro F. With the help of these measures, we have analyzed the results of the classifiers and compared SVM results with Abu Jbara et al. [12] and Hernandez-Alvarez [19]. These approaches were found more relevant to our citations' functions and data set as compared to the other approaches. Our proposed approach achieved 90% macro F as compared to the results of Abu Jbara et al. [12] and Hernandez-Alvarez [19] whom macro F was 58% and 87% respectively for performing Citations' Functions classification. The comparison of results are shown in Figure 5.6.

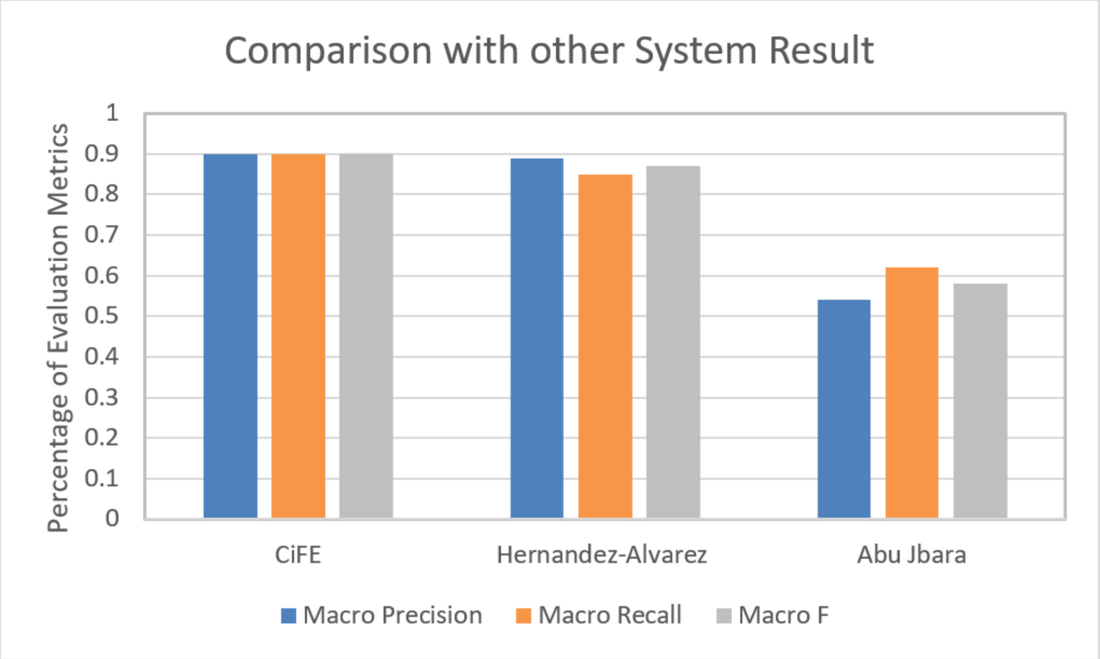

Figure 5.6: Comparison with Other System Results

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Digital libraries have taken the central place in this era of science and technology. The digital libraries contain a huge amount of research articles which is accessible on-line. As the number of research articles are growing, the scholars are now taking interest in analyzing the content of research articles, specifically the context of citation. Understanding the authors' thoughts or sentiments in a large number of research articles published in different areas on daily basis in digital libraries is very helpful to improve citation classification methods. The online libraries enable researchers to find information about the publications and their citations but these libraries do not tell the reason behind citing a paper. Citation analysis techniques are mostly concerned with citation counts. Their applications have led to criticism about simple counts quality. Many researchers have conducted the experiments to analyze the context of the citations for more in-depth insights into scientific knowledge rather than simple counts and have listed various citations' functions. We reviewed the literature in citation classification various categorization schemes have been closely examined, and their automatic classification experiments combined with the machine learning algorithms are also well studied. A detailed study shows that the literature contains more than 150 functions for defining a citation relationship among articles. Annotating every citation into 150 functions is probably impossible. In addition, the functions gathered for the citation often have overlapped as well as diffused meanings. These functions for the citations are discovered by using machine learning algorithms, there will be very low accurate results. To overcome this issue, we have adopted a well-defined CCRO ontology classes for citations' functions classification. There are eight citations' functions in a list which defines a unique citation relation among the research articles. This minimum set of citations' functions defines CCRO classes, which allow Machine learning algorithms to identify the functions for those citations accurately.

In our experiments, we have used supervised machine learning approaches which require annotated data set. We have utilized a specific version of the AAN data set. Our data set is based on Athar's data set. We have randomly selected a set of three hundred sentences from this data set. Manual annotation approach can be used for citing sentences. Annotated data set helped to train the machine learning classifiers to classify the new citing sentences into eight citations' functions. After the completion of annotation process, we apply pre-processing techniques of data mining on annotated data set which transforms the data set into a comprehensible format. We have implemented different steps for pre-processing. For example; Tokenization, Noise Removal, Stop Word's Removal, Lemmatization and POS Tagging. In the classification of citation, the selection of features is an important technique. We have conducted several experiments after the completion of preprocessing steps to identify the important features for accurate classification. We have trained the NB, RF and SVM classifiers on nouns, verbs, adverbs and adjectives. These experiments helped us a lot in the process of devising feature selection rules. We have devised rules with the help of these important features and underlined words or phrases by annotators. Several types of features that capture the characteristics of citation sentences are extracted by devised feature extraction rules are served as the inputs of automatic classifiers. Along with this, the words and phrases which were underlined during annotation process also helped in devising rules. After that, we have trained our classifiers NB, RF and SVM on important extracted features. We have used 10-fold cross validation approach to analyze the results of these classifiers. In this way, we achieved 91% weighted-average F1-score. These results reveals that in classification while using feature extraction rules have outperformed the results which were not using the rules. Furthermore, result of SVM classifier has outperformed other classifiers. We have used these feature extraction rules and trained SVM classifier for the classification of Athar's data set into 8 citations' functions. We have divided 8428 sentences of this data set into 17 sub sets. Every sub set is consisted on 495 sentences. We predicted all citations in 17 steps by using SVM classifier based on the principle of train and test. During the prediction process, we picked up 10% data randomly at first 5 prediction steps. We checked the results and found that 90% of the predicted sentences were correct on average. After the completion of prediction, 15% of citing sentences were randomly picked from the predicted data. We have assigned these sentences to English Linguistic annotator and asked them to annotate these sentences into 8 citations' functions. After manually annotating of 15% citing sentences, we have calculated inter annotator agreement between annotator and machine predictions. For evaluation of inter annotator agreement, we have used the Cohen Kappa Coefficient to evaluate the agreement. The agreement was calculated Kappa 0.737.

After the calculation of inter annotator agreement we used 10-fold cross validation mode for evaluation of classification. The proposed approach attained 98% weighted-average F1-score. The experiments showed that the result of SVM classifier has outperformed other classifiers. We have also compared our results of SVM classifier with Abu Jbara et al. [12] and Hernandez-Alvarez [19]. Our proposed approach achieved good results as compare to Abu Jbara et al. [12] and Hernandez-Alvarez [19] for performing Citations' Functions classification. This classifier is useable in digital libraries to categorize the cited articles into eight citations' functions accurately. The categorization of cited paper in eight citations' functions facilitates the researcher to get understanding of cited paper even before and without reading that paper. With the help of this proposed systems, scholarly community will be able to find maximum number of relevant research papers within minimum time span unlike traditional methods.

6.2 Future work

We have identified some research gap which could be addressed in future. These research gaps are described below:

1. The output of this research can be used in modern digital libraries to categorize the cited articles into 8 citations' functions.
2. As we checked 1314 citation texts from 8428 citation texts, which was 15% predicted data, the remaining predicted data set can be checked in future.

Bibliography

[1] . D. K. S. Voos, H., "Are all citations equal? or, did we op. cit. your idem?." *Journal of Academic Librarianship*, vol. 1, no. 6, pp. 19–21, 1976.

[2] E. Garfield *et al.*, "Can citation indexing be automated," *Statistical association methods for mechanized documentation, symposium proceedings*, vol. 269, no. 2, p. 189–192, 1964.

[3] B.-A. Lipetz, "Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators," *American documentation*, vol. 16, no. 2, pp. 81–90, 1965.

[4] A. Athar, "Sentiment analysis of citations using sentence structure-based features," *Proceedings of the ACL 2011 student session*, vol. 26, no. 7, pp. 81–87, 2011.

[5] E. Orduña-Malea, J. M. Ayllón, A. Martín-Martín, and E. D. López-Cózar, "About the size of google scholar: playing the numbers," *arXiv preprint arXiv:1407.6239*, vol. 8, no. 5, pp. 11–20, 2014.

[6] I. Ihsan and M. A. Qadir, "Ccro: Citation's context and reasons ontology," *IEEE Access*, vol. 7, no. 12, pp. 30 423–30 436, 2019.

[7] M. J. Moravcsik and P. Murugesan, "Some results on the function and quality of citations," *Social studies of science*, vol. 5, no. 1, pp. 86–92, 1975.

[8] S. Teufel, A. Siddharthan, and D. Tidhar, "Automatic classification of citation function," *Proceedings of the 2006 conference on empirical methods in natural language processing*, vol. 6, no. 12, pp. 103–110, 2006.

[9] R. E. Mercer and C. Di Marco, “The importance of fine-grained cue phrases in scientific citations,” *Conference of the Canadian Society for Computational Studies of Intelligence*, vol. 6, no. 12, pp. 550–556, 2003.

[10] C. Dong and U. Schäfer, “Ensemble-style self-training on citation classification,” *Proceedings of 5th international joint conference on natural language processing*, vol. 15, no. 5, pp. 623–631, 2011.

[11] B. H. Butt, M. Rafi, A. Jamal, R. S. U. Rehman, S. M. Z. Alam, and M. B. Alam, “Classification of research citations (crc),” *arXiv preprint arXiv:1506.08966*, vol. 6, no. 8, pp. 81–90, 2015.

[12] V. Q. R. Jha, A. A. Jbara and D. R. Radev, “Nlp-driven citation analysis for scientometrics,” *Natural Lang. Eng.*, vol. 23, no. 1, pp. 93–130, 2017.

[13] A. Athar and S. Teufel, “Context-enhanced citation sentiment detection,” *Proceedings of the 2012 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies*, vol. 11, no. 9, pp. 597–601, 2012.

[14] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, “The acl anthology network corpus,” *Language Resources and Evaluation*, vol. 47, no. 4, pp. 919–944, 2013.

[15] I. C. Kim and G. R. Thoma, “Automated classification of author’s sentiments in citation using machine learning techniques: A preliminary study,” *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, vol. 11, no. 9, pp. 1–7, 2015.

[16] G. Parthasarathy and D. Tomar, “Sentiment analyzer: Analysis of journal citations from citation databases,” *2014 5th international conference-confluence the next generation information technology summit (confluence)*, vol. 8, no. 15, pp. 923–928, 2014.

[17] C. A. Sula and M. Miller, “Citations, contexts, and humanistic discourse: Toward automatic extraction and classification,” *Literary and Linguistic Computing*, vol. 29, no. 3, pp. 452–464, 2014.

[18] J. Xu, Y. Zhang, Y. Wu, J. Wang, X. Dong, and H. Xu, “Citation sentiment analysis in clinical trial papers,” vol. 2015, no. 9, pp. 13–34, 2015.

[19] M. Hernández-Alvarez and J. M. Gómez, “Citation impact categorization: for scientific literature,” *2015 IEEE 18th International Conference on Computational Science and Engineering*, vol. 6, no. 4, pp. 307–313, 2015.

[20] Z. Ma, J. Nam, and K. Weihe, “Improve sentiment analysis of citations with author modelling,” *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, vol. 8, no. 14, pp. 122–127, 2016.

[21] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, “The acl anthology network corpus,” *Language Resources and Evaluation*, vol. 47, no. 4, pp. 919–944, 2013.

[22] S. Teufel, A. Siddharthan, and D. Tidhar, “Automatic classification of citation function,” *Proceedings of the 2006 conference on empirical methods in natural language processing*, vol. 18, no. 9, pp. 103–110, 2006.

[23] J. Cohen, “Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit.” *Psychological bulletin*, vol. 70, no. 4, p. 213, 1968.

[24] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *biometrics*, vol. 18, no. 9, pp. 159–174, 1977.

[25] M. Honnibal and I. Montani, “Spacy 2: natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing,” *To appear*, vol. 7, no. 1, pp. 411–420, 2017.

[26] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," *arXiv preprint cs/0205070*, vol. 27, no. 11, pp. 511–520, 2002.

[27] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," vol. 25, no. 8, pp. 347–354, 2005.

[28] P. D. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," *arXiv preprint cs/0212032*, vol. 18, no. 9, pp. 411–420, 2002.

[29] J. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," *International conference on intelligent text processing and computational linguistics*, vol. 28, no. 9, pp. 486–497, 2005.

[30] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," *Computational linguistics*, vol. 35, no. 3, pp. 399–433, 2009.

[31] . L. John, G.H., "Estimating continuous distributions in bayesian classifiers," *UAI*, vol. 2, no. 4, pp. 8–12, 1995.

[32] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 2, pp. 5–32, 2004.

[33] C. K. H. C. W. X. . L. C. Fan, R., "Liblinear: A library for large linear classification," *Journal of machine learning research*, vol. 9, no. 8, pp. 1871–1874, 2008.

[34] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.